**AP STAT: CHAPTER 3 CATEGORICAL DATA** MAKE A PICTURE!

**First, create a frequency table**

Example: number of students at CB South in each grade:



Proportion = Percent =

Frequency = Relative frequency =

Distribution (of a variable)-

Examples:

**Categorical Distributions:**

1. **Bar Chart**

1. **Pie Chart**
2. **Contingency tables (aka 2-Way tables)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Frosh** | **Soph** | **Junior** | **Senior** | Total |
| **Male** |  |  |  |  |  |
| **Female** |  |  |  |  |  |
| Total |  |  |  |  |  |

Identify:

* Row variable
* Column variable
* Values of the variable
* Total (n)
* # of Cells
* Totals

***Example:*** Hospitals

|  |  |  |
| --- | --- | --- |
|  | **Hospital A** | **Hospital B** |
| **Died** | 63 | 16 |
| **Survived** | 2037 | 784 |

* What percent of people died?

**Notation:**

Probability: Given/Of: And: Or:

* **Of** those people that went to Hospital A, what percent died?
* **Given that** someone went to Hospital B, what is the chance that they died?
* **Of** those people who died, what percent went to Hospital A?
* What percent of people died **and** went to Hospital B?
* What percent of people survived **or** went to Hospital A?

***2 types of Distributions for Categorical Variables***

1. **MARGINAL DISTRIBUTIONS**

* How to make:
* Example: Hair color vs. Gender

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Brown** | **Blonde** | **Black** | **Red** | **Total** |
| **MALE** | 26 | 24 | 10 | 3 | 63 |
| **FEMALE** | 20 | 35 | 12 | 6 | 73 |
| **TOTALs** | 46 | 59 | 22 | 9 | 136 |

* Find the marginal distribution for the HAIR COLOR variable
* Find the marginal distribution for the GENDER variable
* MAKE A PICTURE! BAR CHART

1. **CONDITIONAL DISTRIBUTIONS**

* Look at …
* Then look at …

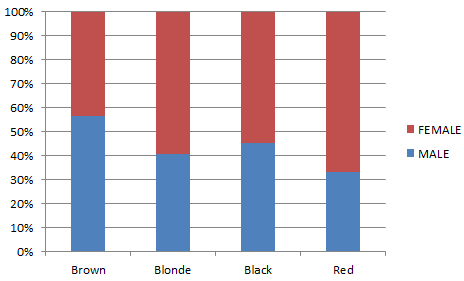
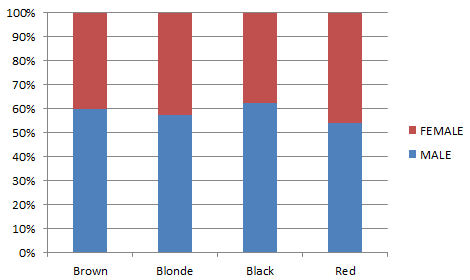
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Brown** | **Blonde** | **Black** | **Red** | **Total** |
| **MALE** | 26 | 24 | 10 | 3 | 63 |
| **FEMALE** | 20 | 35 | 12 | 6 | 73 |
| **TOTALs** | 46 | 59 | 22 | 9 | 136 |

* Break down …
* ALWAYS …
* Example: Hair Color vs. Gender
* Find the conditional Distribution for the HAIR COLOR variable
* Find the conditional Distribution for the GENDER variable
* Represented visually: SEGMENTED (or STACKED) BAR GRAPH
  + Each bar = 100%
  + Values of variable on the x-axis
  + Bars are segmented into parts of each value

**Independence:**

**How do we tell independence?**

**Independent: Dependent:**



**AP Stat- worksheet 3A- Categorical Variables practice**

In a survey of adult Americans, people were asked to indicate their **age** and to categorize their **political preference** (liberal, moderate, conservative). The results are as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Liberal | Moderate | Conservative | *Total* |
| under 30 | 83 | 140 | 73 | *296* |
| 30 - 50 | 119 | 280 | 161 | *560* |
| over 50 | 88 | 284 | 214 | *586* |
| *total* | *290* | *704* | *448* | *1442* |

1. What are the row and column variables?
2. What percent of Liberals are under 30?
3. Of those over 50, what percent are Liberals?
4. Of those that are moderates, what percent are 30-50?
5. What percent of respondents are moderate and under 30?
6. Calculate the ***marginal distribution*** for the ***AGE variable***. Write these down. Then make a bar graph of the marginal distribution for age.
7. Calculate the ***marginal distribution*** for the ***PREFERENCE variable***. Write these down. Then make a bar graph of this marginal distribution.
8. Calculate the ***conditional distribution*** of the ***AGE variable***. Write these down. Then make a segmented bar graph of this marginal distribution.
9. Calculate the ***conditional distribution*** of the ***PREFERENCE variable***. Write these down. Then make a segmented bar graph of this marginal distribution.
10. Are the two variables independent?

**AP Stat- worksheet 3B- Categorical Variable practice**

A 4-year study reported in *The New York Times*, on men more than 70 years old analyzed blood cholesterol and noted how many men with different cholesterol levels suffered nonfatal or fatal heart attacks.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low cholesterol | Medium cholesterol | High cholesterol |
| Nonfatal heart attacks | 29 | 17 | 18 |
| Fatal heart attacks | 19 | 20 | 9 |

1. Calculate the marginal distribution for cholesterol level and make a bar graph.
2. Calculate the marginal distribution for severity of heart attack and make a bar graph.
3. Calculate three conditional distributions for the three levels of cholesterol and make a stacked bar graph.
4. Calculate the conditional distributions for the type of heart attack and make a stacked bar graph.
5. Are the two variables independent?